View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 02-33
Economics Series 14
February 2003

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 98 75

ASSESSING THE EFFECTOS OF MEASUREMENT ERRORS ON THE ESTIMATION OF THE PRODUCTION FUNCTION *

Carmine Ornaghi[1]

Abstract

This article explores the reasons why GMM estimators of production function parameters are generally found to produce unsatisfactory results. I attribute this finding to the inaccurate construction of the variables used in production function analysis. In particular, I suggest that the problem lies in the use of common price deflators as well as in a capital variable that does not reflect the true flow of capital services because of short-run equilibrium effects. I show that the practice of using industry-wide deflators leads to lower scale estimates, mainly due to a relevant downward bias in the labour coefficient. At the same time, it introduces a large upward bias in estimating the elasticity of output with respect to technological innovation. Moreover, a significant improvement in the estimates of capital coefficients is obtained if the information on the degree of capacity utilization is adequately exploited.

Keywords: Scale economies; Price deflators; Temporary Equilibrium.

JEL Classification: C23, L6.

[1] Departamento de Economía, Universidad Carlos III de Madrid. E.mail: carmine@eco.uc3m.es

* I wish to thank Jordi Jaumandreu and Pedro Marin for many helpful discussions and suggestions.

1 Introduction

Despite the wide literature available dealing with empirical estimation of production functions, there are some interesting and challenging questions that have not been solved. Although simple *OLS* regression gives plausible parameter estimates, consistent with constant return to scale, econometricians agree that the usual exogeneity assumptions of the regressors that are required for the consistency of *OLS* are unlikely to hold. To the extent that a productivity shock is anticipated before the optimal quantity of inputs is chosen, the production function disturbances are transmitted to the input demand equation. This can lead to an estimation bias due to the correlation between the right-hand variables and the error term (often defined as "simultaneity bias"). Moreover, it is likely that each firm is characterised by specific factors of production, such as entrepreneurial ability, that are not observable but that can affect the current input choice, thus worsening the bias in *OLS* estimates.

The availability of company panel data provides the necessary set of responses to get around the strict exogeneity requirements of the regressors. As long as one is willing to assume that productivity differences among firms tend to be rather persistent over time, one can proceed to first difference the panel data to eliminate firm-specific effects. At the same time, panel data provides the necessary instruments to correct for simultaneity in this first differenced equations. The most widely used technique is the *Generalized Method of Moment* (*GMM*) that relies on combining in an optimal way a set of orthogonal conditions generally defined using suitably lagged levels of the inputs as instruments.[1]

In empirical practice, attempts to control for unobserved heterogeneity and simultaneity generally give less satisfactory results: low and often insignificant capital coefficients and unreasonably low estimates of returns to scale. A number of reasons has been put forward to explain the endemic problems found with estimators in differences. Blundell and Bond (2000) suggest that these problems derive from the weak correlation that exists between the growth rates of capital and employment and the lagged levels of these inputs. Being the capital and the employment series highly persistent,

[1] See Arellano and Bond (1991) for detailed discussion.

past levels of these variables contain little information about the specification in first differences. They show that estimation can be notably improved by adding lagged first differences as instruments for the equation in levels. Using this extended *GMM* estimator, they find a higher and significant coefficient for the capital and they fail to reject the constant return to scale hypothesis. The explanation and proposed solution both lie in the econometric field.

On the contrary, Klette and Griliches (1996) - hereafter KG - suggest that the implausible low estimates of returns to scale can be explained by the incorrect measure of the output. They affirm that "*the practise of using deflated sales as a proxy for real output will, ceteris paribus, tend to create a downward bias in the scale estimate obtained from the production function regressions*" (pag. 344). They stress that the estimated parameters in production function analysis are reduced form parameters that derive from the interaction of the production function with the demand equation. True elasticities of scale can only be inferred modelling a more general framework where the production function is augmented with a demand equation and a price setting rule. The authors present also an empirical estimation to illustrate their theoretical approach. Unfortunately, without the availability of firm level data, it is impossible to know the magnitude of the bias in scale estimate due to the use of common price deflators and we can not contrast whether reinterpreting the parameters as reduced form coefficients actually provides a reliable solution to this bias.

The first aim of this paper is to provide interesting evidences on this matter. The data base used reports variations in output prices at the firm level. This allow us to construct two measures of output: first deflating sales using firm-level price data (the correct measure of output) and second using industry-wide deflators (the incorrect but generally used measure of output). Apart from investigating the impact on input coefficients for a standard production function, I explore the consequences of using deflated revenues when estimating price-cost margins and scale economies. The results obtained support the perspective of KG. I find that scale estimates and price-cost margins are downward biased when using deflated sales (or deflated value added) as dependent variable in the production function equation.

Additionally, evidences are provided on two aspects that have been largely unnoticed in this literature. First, the fact that the difficulty pointed out for deflated sales needs to be extended to intermediate consumptions. Given that most of the available data set do not report a "quantity" measure of material, the general approach is to deflate the normal expenditure in intermediate inputs by a common (and often unique) deflator. The underlying hypothesis made is that there is perfect competition in this factor market, so that all the firms are charged with the same price. However, significant dispersion in input prices seems to exist between firms in the same industry.[2] The theoretical part of the paper shows that the use of deflated expenditures in materials exacerbates the bias in the point estimates of labour (already) introduced by the use of deflated sales but it can potentially offset the bias in estimating the coefficient for materials. Using the available data on intermediate input prices for individual firms, I present evidences supporting this result. Second, the consequences of using incorrect measure of output and intermediate inputs on estimating R&D output elasticities. Since the seminal paper of Griliches (1979), many authors have investigated the productivity effects of R&D (generally defined as knowledge capital). I find that the practice of using common deflators tend to create a large upward bias in the estimation of the R&D coefficient.

Another reason that has been advanced to justify the poor results obtained in the first difference *GMM* framework is the lack of information on quality of labour and capital and, even more important, on capacity utilisation. As long as demand shocks or sudden changes in factor prices, such as the energy price shocks of 1973 and 1979, lead to under or over-utilisation of capacity, producers may find themselves in short-run or temporary equilibrium. As suggested by Berndt and Fuss (1986) (hereafter BF), in these circumstances the growth rate of output can hardly be explained by the variation of capital as reported in standard company data set. It can be the case that increases in the output produced in two subsequent periods can be accommodated through changes in the utilisation rate of equipment without any further investment in capital. The existence of a temporary equilibrium requires the adjustment of the (quasi-fixed) capital variable in order to get a

[2]Differences in the prices of materials paid by manufacturers can be due to differences in firms' purchasing power or because of incomplete information on the prices charged by all potential suppliers. Large firms, for example, are likely to get lower prices because they can afford higher searching costs (when chosing among suppliers) and because of the discounts that are usually granted on large buying. Beaulieu and Mattey (1999) find evidences of large dispertion in the prices of materials for same US industries.

correct measure of the capital coefficient in production function regressions.

The second goal of this paper is to explore the possibility of improving the estimation of the capital coefficient by taking into account the short-run relative factor usage. Although the relevance of these temporary equilibrium effects has long been recognised in productivity analysis and growth theory, this issue has been largely ignored in empirical research, mainly because of the lack of suitable micro data. I show that the use of capacity utilisation data at the firm level can significantly improve the estimation of the capital coefficient, avoiding the low estimates that are usually found in most of the studies.

The empirical part is based on a micro panel data set of Spanish manufacturing firms that includes about 2000 entities during the period 1990-1999. In addition to the standard data on firms' sales and factors of production, there is access to output and intermediate input prices as well as the average percentage of capacity used by the firm along the year. This data set allows me to compute a "true" measure of output and material and to adjust the stock of capital for short-run equilibrium.

The rest of the paper is organised as follows: Section 2 provides a general theoretical analysis of the bias in the production function regression when using deflated sales and deflated expenditure in intermediate inputs. The analysis is extended to the estimation of capital coefficients considering the information on utilisation of capacity. Section 3 describes the data and explains how the variables are constructed. Section 4 discusses the empirical results while Section 5 draws some final comments.

2 Theoretical and Econometric Framework

In this section I provide a set of theoretical underpinnings to clarify the consequences of using incorrect measure of output and factors of production. KG show that if real but unobserved prices are correlated with the explanatory variables included in the model, an omitted variable bias will arise. The same analysis needs to be extended to intermediate input prices: using common deflators for expenditures in materials can exacerbate or reduce the bias mentioned above. Moreover, plausible assumptions suggest that the R&D coefficient is upward biased. In the second part of the section, I discuss the importance of modifying the standard approach used to estimate the capital coefficient in order to accommodate forms of temporary equilibrium.

2.1 Deflators Bias

To keep things as simple as possible, assume that the production function for manufacturing firms can be represented by a Cobb-Douglas function in three "conventional" inputs,[3] labour L, physical capital C and materials M, augmented with a technology parameter based on the firm research effort R:

$$Q_{it} = L_{it}^{\alpha_1} M_{it}^{\alpha_2} C_{it}^{\alpha_3} R_{it}^{\alpha_4} e^{u_{it}^p} \tag{1}$$

where Q is the quantity produced and u^p is the random error term for the equation, representing the effect of efficiency differences, functional form discrepancies and measurement errors. Subscripts are reported only when it is strictly necessary to avoid confusion.

Taking logarithms and first differences to eliminate fixed effects that are buried in the residual, we obtain the linear equation:

$$\tilde{q}_{it} = \alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{m}_{it} + \alpha_3 \tilde{c}_{it} + \alpha_4 \tilde{r}_{it} + \tilde{u}_{it}^p \tag{2}$$

Where lowercase letter with a tilde stands for logarithm differences between year $t-1$ and t.

Assume that deflated sales is the available proxy for real output. Then, the relationship between output growth, $\tilde{q}$, and deflated sales, $\tilde{y}$, is:

$$\tilde{q}_{it} + \tilde{p}_{it} - \tilde{p}_{It} = \tilde{y}_{it} \tag{3}$$

where $\tilde{p}_{it}$ is the growth in firm i's specific price while $\tilde{p}_{It}$ is the growth in the industry-wide deflator.[4]

[3] As stressed by Griliches and Mairesse (1984, pag. 342), *"one could, of course, consider more complicated functional forms, such as the CES or Translog functions. [..] this will not matter as far as our main purpose of estimating the output elasticities of R&D and physical capital, or at least their relative importance, is concerned.*

[4] Suppose that we have data on sales, $P_{it} * Q_{it}$ and we use a common price deflators P_{It} to get deflated sales $Y_{it} = (P_{it} * Q_{it})/P_{It}$. Taking logarithms and first differences, we obtain $\tilde{y}_{it} = \tilde{q}_{it} + \tilde{p}_{it} - \tilde{p}_{It}$.

Substituting (3) in equation (2), we get:

$$\tilde{y}_{it} = \alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{m}_{it} + \alpha_3 \tilde{c}_{it} + \alpha_4 \tilde{r}_{it} + (\tilde{p}_{it} - \tilde{p}_{It}) + \tilde{u}^p_{it} \quad (4)$$

The latter equation shows that when deflated sales, Y, replace the true measure of output, Q, the omitted price variable $(\tilde{p}_{it} - \tilde{p}_{It})$ enters into the residual. As long as the optimal choice of factors of production is affected by changes in firm prices, the omitted price term tends to create a bias in *OLS* estimator of the parameter vectors α.[5]

At the same time, if the variable that we observe is not real materials, M, but deflated expenditures in materials, N, then we have that $\tilde{n}_{it} = \tilde{m}_{it} + \tilde{g}_{it} - \tilde{g}_{It}$, where $\tilde{g}_{it}$ and $\tilde{g}_{It}$ are intermediate input prices at firm level and aggregate level, respectively. Accordingly, equation (4) should be restated as follows:

$$\tilde{y}_{it} = \alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{n}_{it} + \alpha_3 \tilde{c}_{it} + \alpha_4 \tilde{r}_{it} + (\tilde{p}_{it} - \tilde{p}_{It}) - \alpha_2(\tilde{g}_{it} - \tilde{g}_{It}) + \tilde{u}^p_{it} \quad (5)$$

The term $(\tilde{g}_{it} - \tilde{g}_{It})$ can possibly exacerbate or offset the biased introduced by the output price variable $(\tilde{p}_{it} - \tilde{p}_{It})$.[6] The difficulties outlined above cannot be solved by using appropriate instrumental variables. Potentially useful instruments, which have to be correlated with the inputs growth, are likely to be correlated with the input and output price changes buried in the residual and therefore are illegitimate instruments. Our results, obtained using instrumental variable method confirm this perspective.

Following the analysis presented in KG, the consequences of using common deflators on estimating the production function parameters can be illustrated defining a log-linear demand equation for firm i of the following type:

[5] Let us define the term $(\tilde{p}_{it} - \tilde{p}_{It})$ as Π_{it}. As shown in Klette and Griliches (1996), pag. 346, the bias in the *OLS* estimator of the α coefficients depend on the sign and the magnitude of the parameter δ in the auxiliary regression: $\Pi = X\delta + u^{\Pi}$ where X is the matrix of the growth in inputs and u^{Π} is a white-noise error term.

[6] Let us define the term $(\tilde{g}_{it} - \tilde{g}_{It})$ as Ω_{it}. Using a symmetric argument to the one in the previous footnote, we have that the *OLS* estimator of the parameter vector α depends also on the auxiliary regression: $\Omega = X\gamma + u^{\Omega}$. Then, the overall direction and size of the bias depend on the sign and magnitude of δ and γ.

$$Q_{it} = Q_{It} * (P_{it}/P_{It})^{\eta} R_{it}^{\beta_1} e^{u_{it}^d} \tag{6}$$

where Q_{it} stands for firm i's sales, Q_{It} is the industry output, P_{it} is the firm specific price while P_{It} is the average price in the industry. R_{it} is the knowledge capital of the firm measuring the innovative content of its product. This variable acts as a demand shifter: increasing the quality of the product through innovation boasts the sales of the firm.[7] Rewriting the demand system in terms of growth rates (more precisely as first differences of logarithms), we obtain:

$$\tilde{q}_{it} = \tilde{q}_{It} + \eta(\tilde{p}_{it} - \tilde{p}_{It}) + \beta_1 \tilde{r}_{it} + \tilde{u}_{it}^d \tag{7}$$

Now, using equation (3) and rearranging, it follows that:

$$(\tilde{p}_{it} - \tilde{p}_{It}) = \frac{1}{1+\eta}(\tilde{y}_{it} - \tilde{q}_{It} - \beta_1 \tilde{r}_{it} - \tilde{u}_{it}^d) \tag{8}$$

Equation (8) gives an analytical interpretation of the (usually) unobserved term $(\tilde{p}_{it} - \tilde{p}_{It})$. Apart from the residual, the only variable that is unknown is the growth in industry output $\tilde{q}_{It}$. As shown by KG, it is possible to use the weighted average growth in deflated sales of all the firms in the sample as a proxy for this term . We will come back to this point in Section 3.

Combining equation (5) and (8), we get:

$$\begin{aligned} \tilde{y}_{it} &= \frac{\eta+1}{\eta}(\alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{n}_{it} + \alpha_3 \tilde{c}_{it}) + (\frac{\eta+1}{\eta}\alpha_4 - \frac{\beta_1}{\eta})\tilde{r}_{it} \\ &\quad - \frac{1}{\eta}\tilde{q}_{It} - \alpha_2 \frac{\eta+1}{\eta}(\tilde{g}_{it} - \tilde{g}_{It}) + \tilde{v}_{it}^d \end{aligned} \tag{9}$$

where $\tilde{v}$ is an error term that captures both demand and productivity shocks.

Equation (9) shows that the estimated parameters for the production function have to be interpreted as reduced form parameters (coming up from supply and demand coefficients). Under a variety of situations, the omitted price variable $(\tilde{p}_{it} - \tilde{p}_{It})$ tends to create a downward bias in scale estimate for

[7] See Klette and Griliches (1996) and Klette (1996) for further details.

OLS technique.[8] As we show in the empirical part, the growth in industry output $\tilde{q}_{It}$ plays a crucial role to identify the true coefficients for the standard inputs of production.

Equation (9) differs from the analysis of KG because of two terms: the knowledge capital, $\tilde{r}_{it}$, and the omitted material price, $(\tilde{g}_{it} - \tilde{g}_{It})$. Notice that the two terms making up the coefficient of $\tilde{r}_{it}$ capture the effect of both product and process innovation. The coefficient $(\frac{\eta+1}{\eta}\alpha_4 - \frac{1}{\eta}\beta_1)$ in equation (9) is greater than α_4 in equation (2) if β_1 is larger than α_4.[9] In a companion paper, Ornaghi (2002), evidence supporting this assumption is provided.[10] We can then infer that the use of common price deflators tend to create an upward bias in the estimation of the productivity effects of R&D. The results presented in Section 4 suggest that this upward bias can be rather large for manufacturing firms: using deflated sales as dependent variable can lead to estimate an R&D coefficient that is almost twice as large as the one we obtain using the "true" output measure.

As far as the term $(\tilde{g}_{it} - \tilde{g}_{It})$ is concerned, it seems plausible that firms that experience higher costs of materials will, *ceteris paribus*, reduced the quantity used of this input in favour of the other short-run factor of production, such as labour.[11] In econometric terms, this is equivalent to assume the existence of a systematic negative correlation of the generally unobserved term $(\tilde{g}_{it} - \tilde{g}_{It})$ with materials and a positive correlation with labour. Given the negative sign in front of the $(\tilde{g}_{it} - \tilde{g}_{It})$ coefficient, the downward bias in the estimate of input elasticities due to the omitted output price $(\tilde{p}_{it} - \tilde{p}_{It})$ is exacerbated in the case of labour but it is (partially or fully) offset for materials.[12] The estimates shown in Section 4 confirm that the coefficient of materials under

[8] The term $(\tilde{p}_{it} - \tilde{p}_{It})$ is a source of downward bias for *OLS* estiamates whenever it is negatively correlated with the explanatory variables in the equation. This seems likely if the firm market power arises mainly from a specific demand for its products. Detailed discussion of this issue can be found in Griliches and Mairesse (1995).

[9] Take the coefficient $(\frac{\eta+1}{\eta}\alpha_4 - \frac{1}{\eta}\beta_1)$ and rewrite β_1 in terms of α_4 as $\beta_1 = n\alpha_4$, with $n \in R^+$. Rearrenging, we obtain that $\frac{1}{\eta}(\eta - (n-1))\alpha_4$. This term is greater than α_4 if $n > 1$, that is if $\beta_1 > \alpha_4$. Moreover, the greater is n, the higher is the coefficient of r_{it} in equation (9) compared to the one in (2).

[10] See Garcia *et all.* (2002) for a similar result.

[11] The assumption made is that these two inputs are to a certain extent substitutes in the production process.

[12] Suppose that we have a correct measure of output, Q. Moreover, we know total expenditure in intermediate inputs but we do not have individual prices charged to the firm

equations (9) - that is, when general deflators are used - and (2) - when the "true" measures of output and materials are used -are not sensibly different from each other, but the labour coefficient is seriously biased downwards.

2.2 Temporary Equilibrium

A well known finding in productivity analysis is the low and often insignificant estimate of capital coefficients when the production function is specified in differences. My claim is that this persistent puzzle is essentially due to the existence of temporary equilibria characterised by under or over utilisation of firms' installed capacity. If the production function is stated in levels, it seems plausible that there is always a positive and significant correlation between the firm output and its available total stock of capital, even when producers are not in a long-run equilibrium. Things are different when we consider a regression in first differences. Increases in the output produced in two subsequent periods can now be accommodated through variations in capacity utilisation of equipment and machinery, without undertaking any investment. Appendix C presents a simple econometric model to show that measurement errors are more likely to severely affects the estimation of the capital coefficient when we go from levels to differences.

The relevance of temporary equilibria, especially those associated with business cycles, has long been recognised in productivity analysis. In their article dated 1986, BF suggest to adjust the quantities of the quasi fixed factors for temporary equilibrium effect. The theoretical underpinnings to this

for these inputs; this means that we can construct a proxy for material using general deflators. Then, the (first-differenced) specification of our production function would be: $\tilde{q}_{it} = \gamma_1 \tilde{l}_{it} + \gamma_2 \tilde{m}_{it} + \gamma_3 \tilde{c}_{it} + \gamma_4 \tilde{r}_{it} + \gamma_5 (\tilde{g}_{it} - \tilde{g}_{It}) + \tilde{u}_{it}$ Let define $(\tilde{g}_{it} - \tilde{g}_{It})$ as Ω_{it}. The *OLS* estimate of the labour and material coefficients would be:

$$\gamma_1 = \frac{cov(\tilde{l}_{it}, \tilde{q}_{it})}{var(\tilde{l}_{it})} - \gamma_5 \frac{cov(\Omega, \tilde{l}_{it})}{var(\tilde{l}_{it})} = \alpha_1 - \gamma_5 \frac{cov(\Omega, \tilde{l}_{it})}{var(\tilde{l}_{it})}$$

$$\gamma_2 = \frac{cov(\tilde{m}_{it}, \tilde{q}_{it})}{var(\tilde{m}_{it})} - \gamma_5 \frac{cov(\Omega, \tilde{m}_{it})}{var(\tilde{m}_{it})} = \alpha_2 - \gamma_5 \frac{cov(\Omega, \tilde{m}_{it})}{var(\tilde{m}_{it})}$$

where α_1 and α_2 are the true coefficients. It is likely that firms that experience a higher growth in prices of intermediate inputs, Ω_{it}, will try to substitute this input with labour. For example, firms can decide to reduce the "outsourcing" of their activities when these contracts became so expensive that the cost of performing it inside is less than contracting it out. This reasoning implies a positive value of the covariance term in the first equation reported above and a negative value for the corresponding term in the second equation. Then, we can expect that the use of general deflators leads to $\gamma_1 < \alpha_1$ and $\gamma_2 > \alpha_2$.

adjustment rely on the fact that the stock of capital should always reflect the unobservable available services from this input. However, the lack of micro data has hampered the possibility of defining adjustment procedures. In this paper I define two variants of the production function equation (2) above in order to deal with temporary equilibrium effects. Detailed explanations of these procedures are postponed to the next Subsection. Results presented in Section 4 confirm that estimates of output elasticities with respect to capital are notably improved by exploiting the information on capacity utilisation rates.

2.3 Input Elasticities, Scale Eonomies and Mark-ups

In this section I state the estimating equations considered in the empirical analysis. All the models are estimated using, on one hand, a "quantity" measure of output, Q, and materials, M (that is, the true but usually not observable measures) and, on the other hand, deflated sales, Y, and deflated expenditure in intermediate inputs, N (the variables generally used in most of the empirical papers). All these models provide interesting evidences on the extent of the bias due to the incorrect measurement of the variables. In the first part, the capital input C correspond to the total stock of equipment as reported by the firm and a measure of capacity utilisation, CU, is added as a (independent) regressor in the production function. In the second part, I provide the alternative specifications dealing with short-run equilibrium effects.[13]

Consider equation (2) above. Under constant return to scale with respect to standard factors of production (labour, materials and physical capital), we have that $\alpha_1 + \alpha_2 + \alpha_3 = \varepsilon = 1$. For interpretive reasons, equation (2) is restated so that deviations from constant returns are measured explicitly:

$$\tilde{q}_{it} - \tilde{c}_{it} = \alpha_1(\tilde{l}_{it} - \tilde{c}_{it}) + \alpha_2(\tilde{m}_{it} - \tilde{c}_{it}) + (\varepsilon^{ob} - 1)\tilde{c}_{it} + \alpha_4 \tilde{r}_{it} + \alpha_5 \widetilde{cu}_{it} + \tilde{u}_{it}^{ob} \quad \text{(S1)}$$

The corresponding specification using deflated variables is:

[13] I prefer to start with a more "conservative" specification based on a standard definition of the capital input, C, in order to clearly disentangle the "omitted price" bias from the effects of temporary equilibrium.

$$\tilde{y}_{it}-\tilde{c}_{it}=\gamma_1(\tilde{l}_{it}-\tilde{c}_{it})+\gamma_2(\tilde{n}_{it}-\tilde{c}_{it})+(\varepsilon^{unob}-1)\tilde{c}_{it}+\gamma_4\tilde{r}_{it}+\gamma_5\widetilde{cu}_{it}+\tilde{u}_{it}^{unob} \quad \text{(S2)}$$

Notice that ε^{ob} and ε^{unob} correspond to the scale elasticity of standard inputs when "quantity" measure of output and materials are observable or unobservable, respectively. Moreover, the error term in (S2) picks up differences between changes in firm output and intermediate input prices and the corresponding industry price indexes, $(\tilde{p}_{it}-\tilde{p}_{It})$ and $(\tilde{g}_{it}-\tilde{g}_{It})$. Section 4 shows estimation results when value added is used as dependent variable, too.

Equation (9) above shows that in principle the growth in industry output, $\tilde{q}_{It}$, ensures identification of demand elasticities and production function parameters. If this were the case, true input elasticities obtained from estimating equation (S1) can be inferred, even when we use deflated gross production and intermediate input expenditure, by adding the term $\tilde{q}_{It}$ to model (S2), as follows:

$$\begin{aligned}\tilde{y}_{it}-\tilde{c}_{it} &= \gamma_1(\tilde{l}_{it}-\tilde{c}_{it})+\gamma_2(\tilde{n}_{it}-\tilde{c}_{it})+(\varepsilon^{unob}-1)\tilde{c}_{it} \\ &\quad +\gamma_4\tilde{r}_{it}+\gamma_5\widetilde{cu}_{it}+\gamma_6\tilde{q}_{It}+\tilde{u}_{it}^{unob} \qquad \text{(S3)}\end{aligned}$$

Results presented in Section 4 are partially consistent with this approach, first modelled by KG.

I have also considered a different specification of the models above. Following the works of Hall (1988) and others,[14] the elasticities of short-run inputs, labour and materials, can be replaced by their income shares corrected for the presence of market power:

$$\alpha_1=\mu*\omega_l \text{ and } \alpha_2=\mu*\omega_m \qquad (10)$$

where μ is the ratio between price and marginal costs, ω_l and ω_m are the (observable) income shares of labour, $\omega_l=(CostLabour/Sales)$, and materials, $\omega_m=(CostMaterials/Sales)$.[15] Under these conditions, the resulting estimating equations are:

[14] See, for example, Klette (1999) .

[15] On an empirical ground, the shares of labour and materials costs in total revenues, ω_l and ω_m, are computed as the averages over adjacent years (i.e., following the Tornquist approximation).

$$\begin{aligned}\tilde{q}_{it} - \tilde{c}_{it} &= \mu_1^{ob}\omega_l(\tilde{l}_{it} - \tilde{c}_{it}) + \mu_2^{ob}\omega_m(\tilde{m}_{it} - \tilde{c}_{it}) + (\varepsilon^{ob} - 1)\tilde{c}_{it} \\ &+\alpha_4\tilde{r}_{it} + \alpha_5\widetilde{cu}_{it} + \tilde{u}_{it}^{ob} \qquad \text{(S4)}\end{aligned}$$

and

$$\begin{aligned}\tilde{y}_{it} - \tilde{c}_{it} &= \mu_1^{unob}\omega_l(\tilde{l}_{it} - \tilde{c}_{it}) + \mu_2^{unob}\omega_m(\tilde{n}_{it} - \tilde{c}_{it}) + (\varepsilon^{unob} - 1)\tilde{c}_{it} \\ &+\gamma_4\tilde{r}_{it} + \gamma_5\widetilde{cu}_{it} + \tilde{u}_{it}^{unob} \qquad \text{(S5)}\end{aligned}$$

These later models allow us to determine the extent of the bias in estimating the mark-up coefficient, μ, as well as the scale parameters, ε, when true measures of the variables, as stated in model (S4), are replaced with variables computed with industry-wide deflators, as in model (S5). Although theoretical models predict the existence of a unique markup term (see equation (10) above), we prefer to estimate μ_1 and μ_2 separately in order to see their variation across specifications. This approach can be used as a simple test to assess the effects of industry-wide deflators: as long as the assumption that $(\tilde{g}_{it} - \tilde{g}_{It})$ intensify the downward bias of the labour coefficient while softening the bias for material holds, we should find a discrepancy between μ_1^{unob} and μ_2^{unob}. If this is the case, μ_2^{unob} should be considered a better indicator of the market power exercised by firms in an industry. I come back to this point in Section 4, after presenting the results for specification (S4) and (S5).

As for specification (S3), the variable $\tilde{q}_{It}$ can be added to (S5) in order to retrieve true mark-up factors when industry-wide deflators are used:

$$\begin{aligned}\tilde{y}_{it} - \tilde{c}_{it} &= \mu_1^{unob}\omega_l(\tilde{l}_{it} - \tilde{c}_{it}) + \mu_2^{unob}\omega_m(\tilde{n}_{it} - \tilde{c}_{it}) + (\varepsilon^{unob} - 1)\tilde{c}_{it} \\ &+\gamma_4\tilde{r}_{it} + \gamma_5\widetilde{cu}_{it} + \gamma_6\tilde{q}_{It} + \tilde{u}_{it}^{unob} \qquad \text{(S6)}\end{aligned}$$

Finally, I present the estimating equations when allowing for the existence of temporary equilibria. As advanced in Section 2.2, two different procedures are defined in order to account for these short-run effects. First, I separate the estimates of the capital coefficient depending on the degree of capacity utilisation reported by the firm:

$$\tilde{q}_{it} = \alpha_1\tilde{l}_{it} + \alpha_2\tilde{m}_{it} + \alpha_3^h D_1\tilde{c}_{it} + \alpha_3^l D_2\tilde{c}_{it} + \alpha_4\tilde{r}_{it} + \alpha_5\widetilde{cu}_{it} + \tilde{u}_{it}^p \qquad \text{(S7)}$$

where D_1 is a dummy that takes value 1 if the capacity utilisation in both year *t* and *t+1* is greater than 80% while D_2 takes value 1 if this condition is not satisfied.[16] We expect the parameter α_3^h to be a better estimate of "true" capital coefficient as the proportionality assumption used to identify capital flows from stocks does not hold when factor usage is rather low.[17] As a second approach to the estimation of a more reliable capital coefficient, I define a "short-run" capital input, *C**, by multiplying *C* and *CU*. Then we estimate the equation:

$$\tilde{q}_{it} = \alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{m}_{it} + \alpha_3 \tilde{c}_{it} + \alpha_4 \tilde{r}_{it} + \tilde{u}_{it}^p \tag{S8}$$

using C^* as instrument of the quasi-fixed factor *C*.[18] The econometric underpinnings of specification (S8) is found in the context of errors in variables. The capital stock C is an erroneous measure of true capital services when capacity utilisation is not full.. I then use C^* as instrument to soften the measurement error. Notice that the last two specifications are defined so as to get a direct estimate of the capital coefficient, α_3, and not to determine deviation from constant return to scale, $(\varepsilon - 1)$, as in previous equations.

2.4 Other Econometric Issues

As mentioned in the introductory section, the required predeterminacy of the right-hand variables to get consistent OLS estimates is unlikely to hold for

[16]As explained in Appendix C, the downward bias in the estimate of the capital coefficient is minimized when we use high values of capacity utilization for two consecutive years. I opt for a threshold value of 80% since it is the average capacity utilisation reported by the firm as shown in Appendix A. Results are robust to alternative specifications (higher than 80%) of this value althought coefficients are less precisely estimated.

[17]On this point BK (pag.11) affirm that "*when economic utilization is full, flows from quasi-fixed inputs can be assumed to be proportional to the stocks [...]. This leads to the replacement of unobserved service flows [...] by observed stocks*"

[18]I could alternatively use C^* as regressor in equation (S8) but I think that specification (S8) allows to differenciate those firms that use the same amount of capital services but whose stock of capital is different. For example, suppose that two competing firms are using two equipments each but one of these firms is not using a third equipment. In this context, the stock of capital, *C*, of the latter is greater but because of under utilisation, the short-run capital services, *C**, that the two firms are actually using coincide. The capital inputs must be differentiated because one firm has capital services readily available without affording any installation cost. Moreover, the capacity utilisation is an average rate for the entire year and it can possibly not reflect the true factore usage in some periods.

some inputs of the production function. In particular, I assume that short-run factors of production, labour and materials, are possibly correlated with the error term. To avoid the "simultaneity bias" due to the transmission of productivity shocks, I use *GMM* estimators with lags of employment and materials as instruments of the endogenous variables. As far as the capacity utilisation is concerned, I test for the endogeneity of this variable in each specification using the "Incremental" Sargan Test.[19] I fail to reject the null hypothesis of strict exogeneity only for models (S4), (S5) and (S6). As shown in Section 4, lagged levels of *CU* are used as instruments when estimating all the other specifications.

Additionally, anecdotal evidence suggests that firms do not adopt a uniform criterion when reporting the R&D expenditures incurred during the year. Therefore, it is likely that the R&D stock (see Section 3 below) is subject to some measurement errors, creating a problem of endogeneity similar to the one due to simultaneity. Again, this problem is tackled using lags of this variable as instruments. In Section 4 I report the exact set of instruments used to estimate the alternative specifications stated above.

3 Data and Variables

The data used in this study are retrieved from the *Encuesta sobre Estrategias Empresariales*, ESEE, (Business Strategy Survey) an unbalanced panel sample of Spanish manufacturing firms published by the *Fundación Empresa Pública* covering the period 1990-1999. The raw dataset consists of 3,151 firms for a total number of 18,680 observations. A "clean" sample is defined according to a set of criteria which are given in Appendix A. The sample employed here consists of all the firms that have been surveyed for at least three years after dropping all the time observations for which the data required to the estimation are not available. Firms are classified in 15 different sectors, which are listed in Appendix A. This Appendix also reports an explanation

[19] This test is defined using the Sargan Test, S, and the relative degrees of freedom, *(df)*, of two separate regressions: one considering capacity utilisation as an exogenous variable, $S_1(df_1)$, and the other as an exogenous variable, $S_2(df_2)$.The difference between the two Sargan Test, $S_1{-}S_2$, is distributed as a χ^2 with degrees of freedom $(df_1 - df_2)$.

of the variables used across the specifications stated in Section 2.3 together with descriptive statistics.[20]

A unique feature of the data set is the availability of information on the price changes set by the firm together with price changes charged to the firm for its intermediate inputs. This allows me to construct a proper measure of output, Q, and materials, M.[21] Subtracting the latter from the first, we also obtain a quantity measure of value added, VA.

Complementary information about price deflators for gross production, expenditures in materials and gross value added are taken from the (Spanish) National Institute of Statistics (*Instituto Nacional de Estadistica - INE*). The series of output deflators for gross production at fifteen-industry level cover all the period 1990-1999 while those for value added are limited to the period 1990-1997. Using these series, two alternative dependent variables are defined: deflated gross production, Y, and deflated value added, $deVA$. Deflated materials, N, are computed using an overall deflator for non-durable goods. These alternative measures of output and intermediate inputs allow us to infer the magnitude of the bias due to the use of industry-wide deflators.

Labour inputs are represented by man hours, taking into account overtime and lost hours.[22] It is likely that this variable is affected by measurement errors due to rounding-off. The number of employees, *EMP*, is then used as instruments of the hours of work when estimating the production function.[23] Capital, C, is computed recursively from an initial estimate (based on book values adjusted to take account of replacement values) and data on firms' investments in equipment goods:

$$C_{it} = (1 - \delta_j) * C_{it-1} + I_{it} \tag{11}$$

[20] Further detail on the dataset can be found in Garcia *et all.* (2002) and Ornaghi (2002).

[21] Suppose that we have data on gross production (at seller price) for two consecutive years: $P_{it} * Q_{it}$ and $P_{it+1} * Q_{it+1}$. Having access to data on price changes, we can express the figures above in terms of the reference year t: $P_{it} * Q_{it}$ and $P_{it} * Q_{it+1}$. At this point if we take log first-difference, we have a measure of the output growth rate, $(\log(Q_{it+1}) - \log(Q_{it}))$, free from price effects.

[22] More precisely, this variable has been constructed using the number of workers adjusted for the "double counting" of R&D employees, times the normal hours plus overtime and minus lost hours.

[23] This is more a formal than a substantial problem. Estimation results obtained using lag valued of L instead of *EMP* gives similar results.

The subscript j states that we use sectorial estimates of the rate of depreciation, δ. Real capital is then obtained using an overall investment deflator (for durable goods). Another interesting feature of this survey is the availability of data on capacity utilisation, CU. As mentioned above, I construct an alternative measure of capital, C^*, to be used as instrument of C in specification (S8). This (short-term) capital input is computed multiplying the original capital input, C, by the reported capacity utilisation, CU.

The R&D capital variable has been computed using standard historical or perpetual inventory method:

$$R_{it} = (1 - \delta) * R_{it-1} + E_{it-1} \tag{12}$$

The knowledge capital in period t depends on the previous period stock of knowledge, suitably depreciated at rate δ, plus investments in R&D of period *t-1*. These investments, E, take into account not only the cost of intramural activities but also payments for outside contracts and imported technology. Equation (11) has been implemented using a depreciation rate of 15%.[24] Given that our data are limited to the period 1990-1999, I need to define a plausible value of the knowledge stock for 1990 before applying the law of motion (11). The pre-sample capital is defined using the data on the history of R&D expenditures during the eighties and nineties provided by the *INE* for our 15 industries. Once computed the firms' average expenditures during the period 1990-1999 and the associated expenditures at industry level using our survey, the individual R&D is assumed to follow the same evolution of total industry investments for each year since the firm has been established;[25] if the firm has been established before 1980, we just consider the expenditure of the eighties..[26]

[24] This is the depreciation rate most frequently used in this type of estimation. In another paper, Ornaghi (2002), I show that estimates of R&D coefficients are robust to alternative specification of the depreciation rate. Hall and Mairesse (1995) obtain similar evidences.

[25] For example, suppose that firm-i average expenditures during the period 1990-1999 amounts to 5% of total industry-j average expenditures for the same period. Firm-i investments for the previous decade are computed applying this percentage to total industry-j R&D investments as reported by the *INE*.

[26] The definition of the initial capital does not seem to be a relevant issue in the case of Spain, considering that the level of R&D investments during the seventies is negligible and the expenditures during the eighties are sensibly lower than those of the following decade. Total R&D investments amount to 46,862 millions of pesetas in 1982 compared to 246,239 in 1990. The average amount of R&D expenditures for the period 1982-1989 is

Finally, the growth rate in industry output from year *t-1* to year t, $\tilde{q}_{It}$, is estimated using a weighted average of growth in deflated sales for all the firms in the industry, with the average output shares in the two years as weights.[27]

4 Regression Results

Table 1 presents the results obtained estimating the production function (S1) and (S2) as well as model (S3) where industry output is added as an extra regressor. The first interesting evidence that obtained comparing columns 1 and 2 is that scale elasticities are lower when we use deflated sales, Y, and expenditure in intermediate inputs, N, instead of output, Q, and materials, M. In particular, the null hypothesis of constant return to scale is rejected at 5% significant level only in specification (S2). This result is consistent with the idea of downward bias in scale estimates advanced by KG. However, looking at the estimated coefficients of short-run inputs, it appears that only the coefficient of labour is downward biased (and largely so) while the coefficient of materials is quite stable across the two specifications. In Section 2, I support the perspective that the downward bias due to the omitted output price can be either exacerbated or offset when there is also an omitted input price buried in the residuals. This result confirms that the practice of using common price deflators do not affect all the inputs coefficients in the same way. Further detail on this point can be found in Appendix B. There is another problem that has been ignored in production function analysis. As suggested by the theoretical framework in Section 2, the impact of knowledge capital on firms' productivity can be overestimated when firm-level price data are not available. The output elasticity with respect to R&D is more than 70% higher in column 2 than in column 1.

INSERT TABLE 1 ABOUT HERE

All these findings are confirmed when we use value added as the dependent variable. In column 4 the "true" measure of value added, VA,[28] is used while

99,370 millions compared to a higher 262,460 for 1990-1992.

[27] This is the same approach used by KG.

[28] This is defined as the difference between "true" output, Q, and materials, M.

in column 5 the dependent variable, $deVA$, is computed using industry-wide deflators.[29] Scale elasticities are more than ten percentage points lower in column 5 and, once again, the null hypothesis of constant return to scale is rejected only for this specification. Labour and R&D coefficients are seriously downward and upward biased, respectively. Column 3 shows the results of estimating equation (S3). The coefficient of growth in industry output is highly significant. Interestingly, adding this industry variable to the model do not affect the estimated coefficients for the other variables. As argued in Section 2, true elasticities of output with respect to the factors of production can be identified using the coefficient of this new term. Simple calculation shows that model (S3) implies a demand elasticity, η, of -4.31. If we adjust the labour coefficient (the only parameter that is largely affected by price deflators) in column 3 for the term $\frac{\eta}{1+\eta}$ (see equation (9) above), we get a new estimate around 0.38, similar to the result reported in column 1.

Table 2 shows the results obtained estimating models (S4), (S5) and (S6). The same findings stressed above are confirmed under these specifications. Scale elasticities are lower when we use price deflators so that the null hypothesis of constant returns to scale have to be rejected. The R&D coefficient is twice as large as the one reported in column 1, confirming a relevant upward bias in the estimation of the productivity effect of technological innovation. The mark-up parameters μ_1^{ob} and μ_2^{ob} suggest the existence of a moderate market power across the industries. Interesting enough, the null hypothesis that these two parameters are not statistically different cannot be rejected. On the contrary, estimates reported in column 2 reveal a relevant downward bias in the mark-up associated to labour, μ_1^{unob}, while the term, μ_2^{unob}, does not show the same pathology.[30] These results show that the separate estimation of μ_1^{unob} and μ_2^{unob} can be considered a sort of test to assess the effect of industry-wide deflators on the mark-up estimates. Moreover, they seem to support the claim that the mark-up coefficient for deflated intermediate consumption, μ_2^{unob}, is more reliable that μ_1^{unob} as a measure of market power.

INSERT TABLE 2 ABOUT HERE

[29] As mentioned in Section 3, I find value added deflators only for the period 1990-1997.

[30] Point estimate of μ_2 is a little bit higher in column 2 than in column 1 as for the models presented in Table 1.

As far as model (S6) is concerned, we can infer that the demand elasticity, η, is around -6.06.[31] If we adjust the estimate of μ_1^{unob} in column 3 for the term $\frac{\eta}{1+\eta}$, we get a new price-cost margin ratio associated to the labour coefficient of 1.10, very close to the result obtained for specification (S4). This result supports the perspective of KG that input coefficients, scale elasticities and price-cost margins have to be considered reduced-form parameters (a mixtures of supply and demand side coefficients) when using common price deflators as in model (S6).

All the results presented in Table 1 and 2 are based on models specified in such a way that deviation from constant return to scale are measured explicitly. Therefore, the well-known problem of small capital coefficient is somehow hidden. Nevertheless, it is simple to notice that the capital coefficients take low values across all the specifications, thus confirming the problem found with estimators in differences.[32] As mentioned in Section 2, these poor results are possibly due to the fact that the available services from this quasi-fixed input are not fully utilised in all the faces of the business cycle. Table 3 presents the results for specification (S7). The capital coefficient for firms with high degree of capacity utilisation, α_3^h, is positive, large and statistically different from zero. On the contrary, the estimated coefficient for firms with low factor usage, α_3^l, takes a negative sign although it is not statistically different from zero. Results reported in Table 4 strengthen the findings above. Using the "short-run" capital C^* as instrument, we get higher and more reasonable estimates of capital coefficient. The point estimate of α_3 in column 1 of Table 4 is more than three times higher than the lower value of 0.02 that can be inferred from column 1 of Table 1.

INSERT TABLE 3 and 4 ABOUT HERE

As the capacity utilization is not a variable usually provided in many data set, in Appendix D I show that similar results can be obtained using other proxies, such as the individual firm productivity (measured as output per employee) or the sample growth rates of production. The results presented above are consistent with our claim that the low and often insignificant capital coefficients usually found in productivity analysis are due to the

[31] This value is slightly higher than -4.31 as reported above but it can be considered in the same order of magnitude.

[32] For example, the estimated values of the capital coefficient, α_3, are around 0.02 and 0.05 in column 1 and 4 of Table 1, respectively.

existence of short-run equilibrium characterised by sub-optimal utilisation of this quasi-fixed input.

5 Conclusions

This paper explores the reasons why *GMM* estimators of production function parameters are generally found to produce unsatisfactory results: low and often insignificant capital coefficient and unreasonably low estimates of returns to scale. I attribute this well-known finding to the inaccurate construction of the variables used in production function analysis. In particular, I suggest that the problem lies in the use of common price deflators as well as in a capital stock that does not reflect the time flow of the services. I show that the practice of using deflated sales and deflated expenditure in materials instead of a "quantity" measure of output and intermediate inputs leads to lower scale estimates, mainly due to a relevant downward bias in the labour coefficient. At the same time, the unavailability of firm-specific price data introduces a large upward bias in estimating the elasticity of output with respect to technological innovation. I propose a simple test to assess the effect of industry wide price deflators based on comparing the two mark-up parameters of specification (S5). Moreover, I show that the KG suggestion of using $\tilde{q}_{It}$ to infer the true values of input coefficient seems to work adequately.

As far as the capital variable is concerned, a dramatic improvement in coefficient estimates is obtained when we take into account the annual variations in capacity utilisation. As described by Doms and Dune (1998), investment is largely a lumpy activity characterised by period of low investment followed by bursts of investments.[33] When we specify our production function in levels, the impossibility of full capital utilisation in a year of intense investment activity does not trigger the estimate of capital coefficients as long as firms move towards their long-run equilibrium during the sample period. On the contrary, going to first-differences can reveal how dramatic can be the incorrect definition of capital inputs since there are a few periods of intense capital growth but these do not necessary coincide with proportional variations in

[33] This finding is confirmed for the case of Spain. Sanchez (2002) finds evidences of investment spikes using the same survey (*ESEE)* of this study.

capital services if an unexpected under utilisation of capacity occurs. I show that if the flow of capital services is adequately measured, there is a notable improvement in the estimation.

An important message that we can get from the set of results reported above is that imposing the constant return to scale restriction even when we find lower estimates of returns to scale is not incorrect. Measurement errors of output and inputs can seriously bias downward the estimates of input coefficients and the constant return hypothesis would not be rejected if it were not for these measurement errors. Blundell and Bond (2000) find that more reasonable parameters estimates can be obtained using a composed or extended GMM estimators, where lagged first differences are considered informative instruments for the endogenous variables in levels. This paper shows that the greatest improvements in estimating the production function parameters can be derived from a better construction of the (dependent and independent) variables more than from a refinement of the econometric techniques. Although this requires the use of detailed data that are usually not available, I show that more reliable results can be obtained using other proxies, in particular for the degree of capacity utilisation.

Appendix A: Data and Variable

A1 Construction of Data Sample

The survey provides data on manufacturing firms with 10 or more employees. When this was designed, all firms with more than 200 employees were required to participate while a representative sample of about 5% of the firms with 200 or less employees was randomly selected. In 1990, the first year of the panel, 715 firms with more than 200 employees were surveyed, which accounts for 68% of all the Spanish firms of this size. Newly established firms have been added every subsequent year to replace the exits due to death and attrition. We start with a sample of 3,151 firms in an unbalanced panel data. The total number of observations is 18,680. We then clean our dataset according to the following criteria:

1) We remove all the observations where the log difference of the R&D variable between two consecutive years exceeds 4 in absolute value. This removes 64 observations.

2) We drop the observations where the log difference of the capital stock variable between two consecutive years exceed 2 in absolute value. This removes other 81 observations.

The subsample we use in our study consists of all the firms that have been surveyed for at least three years. There are 2,430 firms satisfying this condition, for a total number of 16,637 observations. At this point, we remove any observations for which the data required to the estimation are not available. In the tables showing the results of the estimation, we report the exact number of observations making up the final samples.

The original industrial classification reported in the survey is based on 18 sectors. A classification based on 15 sectors is used here since deflators for gross production and were found at this level because of aggregation. These 15 sectors are:

1) Ferrous and non ferrous metals; 2) Non-metallic minerals; 3) Chemical products; 4) Metal products; 5) Industrial and agricultural machinery; 6) Office and data processing machine; 7) Electrical and electronic goods; 8) Vehicles, cars and motors; 9) Other transport equipment; 10) Food and beverages; 11) Textiles, clothing and shoes; 12) Timber and furniture; 13) Paper and printing; 14) Rubber and plastic products; 15) Other manufacturing products.

A2 Description of Variables

Capacity Utilization (U): Yearly average rate of capacity utilization reported by the firms

Deflated Gross Production (Y): Gross production is defined as the sum of sales and the variation of inventories. We deflate the nominal amount using output deflators at fifteen-industry level.

Deflated Expenditure in Materials (N): We deflate total expenditure in intermediate inputs as reported in the data set using a deflator for non-durable goods.

Deflated Value Added (deVA): It has been computed subtracting N from the gross production Y (see above the two definitions).

Employment (EMP): Approximation to the average number of works during the year; it does not consider employees engaged in R&D activities.

Labour (L): Labour consists of the total hours of work. It has been constructed using the number of works, adjusted for the double counting of R&D employees, times the normal hours plus overtime and minus lost hours.

Materials (M): Nominal materials are given by the sum of purchases and external services minus the variation of intermediate inventories. We use firms' specific deflator based on the variation in the cost of raw materials and energy as reported by the firm.

R&D stocks (R): This variable is constructed using the perpetual inventory method, assuming a depreciation rate of 15%, $\rho = 0.15$. Computation is fully explained in Section 3.

Output (Q): Nominal output is defined as the sum of sales and the variation of inventories. We deflate the nominal amount using the firms's specific output price as reported in the data set.

Physical Capital (C): It has been constructed capitalising firms' investments in machinery and equipment and using sectorial rates of depreciation. The capital stock does not include buildings. This variable is taken from Martin and Suarez (1997).

Value Added (VA): It has been computed subtracting the value of materials, *M*, from output, *Q*. As both these variables are based on firm-level prices, we consider *VA* as our "true" measure of production.

Growth rates of the variable[a]; sample period: 1990-1999.

Variable	**Name**	**Mean**	**St. Dev.**
Output	Q	0.0369	0.279
Deflated Gross Prod.	Y	0.0282	0.279
Value Added	VA	0.0505	0.463
Deflated Value Added	$deVA$	0.0018	0.502
Labour	L	0.0015	0.211
Employment	EMP	0.0023	0.194
Materials	M	0.0266	0.382
Deflated Expen. Mater.	N	0.0276	0.383
Physical Capital	C	0.0775	0.248
Capacity Utilization[b]	U	0.80	0.156
R&D stocks	R	0.0366	0.232

Note:

[a]We use the logarithmic differences of the variables.

[b]Variable expressed in Levels

Appendix B: Decomposing the "omitted price" bias

Results reported in this table show the impact on production function parameters of the term $(\tilde{p}_{it} - \tilde{p}_{It})$ and $(\tilde{g}_{it} - \tilde{g}_{It})$, separately. Column 1 reports estimation results using "true" measure of output, Q, and materials, M. In column2, deflated gross production, Y, is used as dependent variable and materials, M, as a regressor (this allows us to isolate the bias due to $(\tilde{p}_{it} - \tilde{p}_{It})$). In column 3, I use deflated expenditure in intermediate inputs, N, and output, Q, as dependent variable (this allows us to isolate the bias due to $(\tilde{g}_{it} - \tilde{g}_{It})$). Finally, column 4 reports the overall effect on the estimation of the parameter vector α when using industry-wide output and input deflators. Notice that results shown in column1 and 4 are equivalent to those reported in Table 1.

Estimation method: **GMM estimates**[a]

Dep.Vbls	Q	Y	Q	Y
Model	1	2	3	4
Labour	0.362***	0.344***	0.318***	0.296***
	(0.063)	(0.061)	(0.065)	(0.062)
Materials	0.541***	0.537***	0.565***	0.570***
	(0.049)	(0.049)	(0.052)	(0.051)
Capital	-0.077	-0.091*	-0.101*	-0.110**
	(0.056)	(0.055)	(0.056)	(0.054)
R&D	0.112**	0.143***	0.166***	0.192***
	(0.055)	(0.053)	(0.054)	(0.050)
Cap.Util	0.286***	0.288***	0.241***	0.238***
	(0.089)	(0.088)	(0.091)	(0.088)
Ind. Dum.	Inc.	Inc.	Inc.	Inc.
Time Dum.	Inc.	Inc.	Inc.	Inc.
Period	1990-99	1990-99	1990-99	1990-99
N. Obs	11,476	11,476	11,476	11,476
Sargan T.	90.48	98.5	87.68	88.86
(df)	(82)	(82)	(82)	(82)
m1	-10.45	-10.54	-10.34	-10.56
m2	0.01	-0.61	-0.16	-0.66

Heteroskedasticity robust standard errors shown in parentheses.

* significant at 10% level; ** significant at 5%; *** significant at 1%.

[a] $IV\dot{s}$: number of workers, physical capital and capacity utilization lagged levels from $t-2$ to $t-4$; materials lagged levels from $t-2$ to $t-3$ and R&D capital lagged levels in $t-2$; (exogenous variables) growth of capital and growth in industry output (for specification in column 3).

Appendix C: Measurement Errors of Physical Capital

In this Appendix I show that the bias of capital coefficients is likely to be amplified when we specify the production function in differences. This analysis mimics the one presented in Chapter 4 of Arellano (2002), to which we refer for detailed explanations. Consider first the model in levels $y_{it} = \alpha x_{it}^{\dagger} + u_{it}$ and suppose that the regressor is measured with an error, so that we actually observe: $x_{it} = x_{it}^{\dagger} + \varepsilon_{it}$ (i). Under this framework, the existence of temporary equilibria characterized by sub-optimal utilisation of the available physical capital can be modelled as a measurement error $\varepsilon_{it} = (1-\theta_{it})x_{it}$ where θ_{it} is the proportion of capital stock x_{it} that is actually used. Then, we can rewrite (i) as $x_{it} = x^{\dagger}_{it} + (1-\theta_{it})x_{it}$ or, rearranging, $x_{it} = (1/\theta_{it})x^{\dagger}_{it}$. Now, the point estimate of the capital coefficient using x_{it} will be given by:

$$\hat{\alpha} = \frac{cov(x_{it},y_{it})}{var(x_{it})} = \frac{cov((1/\theta_{it})x^{\dagger}_{it},y_{it})}{var((1/\theta_{it})x^{\dagger}_{it})} = \frac{(1/\theta_{it})cov(x^{\dagger}_{it},y_{it})}{(1/\theta_{it})^2 var(x^{\dagger}_{it})} = \theta_{it} * \alpha \text{ (ii)}$$

where α is the true value of the capital coefficient if we use $x^{\dagger}_{it}$. In my dataset, firms use in the average 80% of their capacity utilisation, $\bar{\theta} = 0.8$, so that we can expect a point estimate of the capital coefficient that is lower than its true value. Now, I analyse the effect of measurement errors when the model is specified in first differences: $\Delta y_{it} = \alpha \Delta x^{\dagger}_{it} + \Delta u_{it}$. The estimated coefficient is given by: $\frac{cov(\Delta x_{it},\Delta y_{it})}{var(\Delta x_{it})} = \frac{cov((1/\theta_{it})x^{\dagger}_{it}-(1/\theta_{it-1})x^{\dagger}_{it-1},\Delta y_{it})}{var((1/\theta_{it})x^{\dagger}_{it}-(1/\theta_{it-1})x^{\dagger}_{it-1})} =$

$$\frac{cov(\frac{1}{\theta_{it}}x^{\dagger}_{it}-\frac{1}{\theta_{it}}x^{\dagger}_{it-1}+\frac{1}{\theta_{it}}x^{\dagger}_{it-1}-\frac{1}{\theta_{it-1}}x^{\dagger}_{it-1},\Delta y_{it})}{var(\frac{1}{\theta_{it}}x^{\dagger}_{it}-\frac{1}{\theta_{it}}x^{\dagger}_{it-1}+\frac{1}{\theta_{it}}x^{\dagger}_{it-1}-\frac{1}{\theta_{it-1}}x^{\dagger}_{it-1})} = \frac{cov(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it}+(\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1},\Delta y_{it})}{var((\frac{1}{\theta_{it}})\Delta x^{\dagger}_{it}+(\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1})}$$

$$= \frac{cov(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it},\Delta y_{it})+cov((\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1},\Delta y_{it})}{var(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it})+var((\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1})+2cov(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it},(\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1})} \text{ (iii).}$$

If we assume that the covariance term in the denominator and the second term in the numerator are close to zero (given that most of the series of output and inputs are persistent, first differences have low correlation with past levels as noticed by Blundell and Bond (2000)), we can rewrite (iii) (after dividing by $var(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it})$ the denominator and the numerator) as $\frac{\theta_{it}*\alpha}{1+\lambda}$ (iv) where $\lambda = \frac{var((\frac{1}{\theta_{it}}-\frac{1}{\theta_{it-1}})x^{\dagger}_{it-1})}{var(\frac{1}{\theta_{it}}\Delta x^{\dagger}_{it})}$. Comparing (ii) and (iv), we find that the bias of the capital coefficient for first difference equations is higher than the one for levels. In order to minimize this bias we need to find two (consecutive) values of θ_{it} and θ_{it-1} that are both close to 1. In this way, the downward bias in the numerator (as θ_{it} will be close to one) and in the denominator (as the term $(\frac{1}{\theta_{it}} - \frac{1}{\theta_{it-1}})$ will be close to zero) is simultaneously reduced

Appendix D: Temporary Equilibrium

As data on capacity utilization are usually not available, this Appendix shows that similar results to those presented in Table 3 can be obtained using other proxies, namely the individual firm productivity and the "sample" growth rates of production. Let's start considering the first of these two variables.

After computing the worker productivity, *prod,* measured as output per employee, I determine the maximum value, *maxprod*, and the minimum value, *minprod*, of this variable for each firm across all the years surveyed and, then, the following rule is applied to infer the possible capacity utilization:

$CapUtil2_{it} = (100 - 50 * (maxprod - prod_{it})/(maxprod - minprod))$

The variable *CapUtil2* takes values between 50 (in the year of the lowest worker productivity) and 100 (in the year of highest productivity). Then using this variable, I determine two distinct capital coefficients as in specification (S7) of Section 3.2:

$\tilde{q}_{it} = \alpha_1 \tilde{l}_{it} + \alpha_2 \tilde{m}_{it} + \alpha_3^h D_1 \tilde{c}_{it} + \alpha_3^l D_2 \tilde{c}_{it} + \tilde{u}_{it}^p$ (S7b)

where D_1 is a dummy that takes value 1 if the variable *CapUtil2* is greater than 80% in both year t and $t+1$ while D_2 takes value 1 if this condition is not satisfied. Column 1 of Table D1 reports the results obtained using this approach.

The second variable used as a proxy for capacity utilization is the "sample" growth of production. Once computed the average growth in each year of the survey, I consider the 4 years where the growth rate records a greater increase (or a smaller decrease) compared to the previous year and we estimate the capital coefficient for these 4 years separately from the other years. This means that the dummy D_1 in equation (S7b) takes value one for all the observations in these 4 years (namely, 1991, 1994, 1997 and 1998) and D_2 takes value 1 for the observations in the remaining years. Results reported in column 2 show that even under this simple classification, rather different point estimate of the elasticity of output with respect to capital are obtained.

Estimation method: **GMM estimates**

Dep.Vbls	Q		Q	
	1[a]		**2**[a]	
Labour	0.313***	(0.075)	0.274***	(0.073)
Materials	0.623***	(0.045)	0.613***	(0.046)
$D_1 * Capital$	0.155*	(0.085)	0.095**	(0.038)
$D_2 * Capital$	-0.021	(0.029)	-0.048	(0.035)
Ind. Dum.	Inc.		Inc.	
Time Dum.	Inc.		Inc.	
Period	1990-99		1990-97	
N. Obs	11,581		11,581	
Sargan T. (df)	56.83	(54)	54.8	(54)
m1	-9.46		-9.34	
m2	-0.09		-0.21	

* significant at 10% level; ** significant at 5%; *** significant at 1%.

[a] $IV\dot{s}$: number of workers and physical capital lagged levels from $t-2$ to $t-4$; materials lagged levels from $t-2$ to $t-3$; (exogenous variables) growth of capital.

References

[1] Arellano, M. and Bond S. (1991), "Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations" *Review of Economic Studies*, Vol. 58, pp. 277-297.

[2] Beaulieu, J. and Mattey J. (1999), "The Effects of General Inflation and Idiosyncratic Cost Shocks on Within-comodity Price Dispertion: Evidence from Microdata" *Review of Economics and Statistics,* Vol. 81, pp. 205-216.

[3] Berndt, E.R. and Fuss M. A. (1986), "Productivity Measurement with Adjustment for Variations in Capacity Utilization and Other Forms of Temporary Equilibrium" *Journal of Economietrics*, Vol. 33, pp. 7-29.

[4] Blundell, R. and Bond S.R. (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews,* Vol. 19, pp. 321-340.

[5] Cooper, R., Haltiwanger J. and Power L. (1999) "Machine Replacement and the Business Cycle: Lumps and Bumps" *American Economic Review*, Vol. 89, pp. 921-946.

[6] Doms, M and Dunne T. (1998), "Capital Adjustment Patterns in Manufacturing Plants" *Review of Economic Dynamics,* Vol. 1, pp. 409-429.

[7] Garcia, A., Jaumandreu, J. and Rodriguez, C. (2002), "Innovation and jobs: evidence from manufacturing firms". *mimeo.*

[8] Griliches, Z. (1979), "Issues in Assessing the Contribution of R&D to Productivity Growth". *Bell Journal of Economics*, Vol. 10, pp. 92-116.

[9] Griliches, Z. and Mairesse J. (1995), "Production Function: The Search for Indentification". *NBER Working Paper*, No. 5067.

[10] Hall, B.H and Mairesse J. (1995), "Exploring the Relationship between R&D and Productivity in French Manufactuing Firms". *Journal of Economietrics*, Vol. 65, pp. 263-293.

[11] Hall, R.E. (1988), "The Relation between Price and Marginal Cost in U.S. Industry" *Journal of Political Economy*, Vol. 96, pp. 921-947.

[12] Kettle, T.J. (1999), "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data". *The Journal of Industrial Economics*, Vol. 48, pp. 451-476.

[13] Kettle, T.J. (1996), "R&D, Scope Economies, and Plant Performance". *RAND Journal of Economics*, Vol. 27, pp. 502-522.

[14] Kettle, T.J. and Griliches Z (1996), "The inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous". *Journal of Applied Econometrics*, Vol. 11, pp. 343-361.

[15] Martin, A. and Suarez C. (1997), "El Stock de Capital para las Empresas de la Encuesta sobre Estrategias Empresariales". *Documento Interno n.13, Programa de Investigaciones Económicas,* Fundación Empresa Pública

[16] Ornaghi C. (2002) "Spillovers in Product and Process Innovation: Evidence from Spain",*Working Papers 02-32*, Universidad Carlos III de Madrid..

[17] Sanchez R. (2002) "Estimation of a Dynamic Discrete Choice Model of Irreversible Investment", *Working Papers 01-56*, Universidad Carlos III de Madrid.

TABLES

Table 1. Production Fuction Estimates

Estimation method: **GMM estimates**

Dep.Vbls	Q	Y	Y	VA	$deVA$
Model	(S1)[a]	(S2)[a]	(S3)[a]	(S1)[b]	(S2)[b]
Labour	0.362***	0.296***	0.296***	0.775***	0.621***
	(0.063)	(0.062)	(0.062)	(0.125)	(0.137)
Materials	0.541***	0.570***	0563***	-	-
	(0.049)	(0.051)	(0.051)		
Capital	-0.077	-0.110**	-0.117**	-0.177	-0.293**
	(0.056)	(0.054)	(0.055)	(0.112)	(0.122)
R&D	0.112**	0.192***	0.190***	0.268***	0.441***
	(0.055)	(0.050)	(0.050)	(0.104)	(0.108)
Cap.Util	0.286***	0.238***	0.257***	0.536***	0.688***
	(0.089)	(0.088)	(0.089)	(0.178)	(0.206)
Ind.Output	-	-	0.232***	-	-
			(0.040)		
Ind. Dum.	Inc.	Inc.	Inc.	Inc.	Inc.
Time Dum.	Inc.	Inc.	Inc.	Inc.	Inc.
Period	1990-99	1990-99	1990-99	1990-97	1990-97
N. Obs	11,476	11,476	11,476	8,687	8,687
Sargan T.	90.48	88.86	94.70	57.44	59.28
(df)	(82)	(82)	(82)	(66)	(66)
m1	-10.45	-10.56	-10.55	-10.45	-10.83
m2	0.01	-0.66	-0.56	-1.09	-1.64

Heteroskedasticity robust standard errors shown in parentheses.

* significant at 10% level; ** significatn at 5%; *** significant at 1%.

[a] $IV\dot{s}$: number of workers, physical capital and capacity utilisation lagged levels from $t-2$ to $t-4$; materials lagged levels from $t-2$ to $t-3$ and R&D capital lagged levels in $t-2$; (exogenous variables) growth of capital and growth in industry output (for specification in column 3).

[b] $IV\dot{s}$: number of workers and capacity utilisation lagged levels from $t-2$ to $t-4$; physical capital from t to $t-4$ (as it is considered exgoneous); R&D capital lagged levels from $t-2$ to $t-3$;

Table 2. Mark-up Estimates

Estimation method: **GMM estimates**

Dep.Vbls	Q	Y	Y
Model	(S4)[a]	(S5)[a]	(S6)[a]
Labour	1.095***	0.929***	0.926***
	(0.113)	(0.118)	(0.118)
Materials	1.082***	1.104***	1.104***
	(0.061)	(0.056)	(0.056)
Capital	-0.072	-0.105**	-0.106**
	(0.048)	(0.047)	(0.048)
R&D	0.070**	0.143***	0.143***
	(0.035)	(0.033)	(0.034)
Cap.Util	0.028**	0.039***	0.039***
	(0.012)	(0.012)	(0.012)
Ind.Output	-	-	0.165***
			(0.036)
Ind. Dum.	Inc.	Inc.	Inc.
Time Dum.	Inc.	Inc.	Inc.
Period	1990-99	1990-99	1990-99
N. Obs	9,969	9,969	9,969
Sargan T.	82.64	78.56	81.65
(df)	(78)	(78)	(78)
m1	-12.71	-13.64	-13.59
m2	-1.08	-2.41	-2.26

* significant at 10% level; ** significatn at 5%; *** significant at 1%.

[a]$IV\dot{s}$: number of workers from $t-2$ to $t-5$; physical capital from t to $t-3$ (as it is considered exgoneous); materials lagged levels from $t-2$ to $t-3$ and R&D capital lagged levels from $t-2$; (exogenous) growth rate of capacity utilisation.

Table 3. Temporary Equilibrium 1

Estimation method: **GMM estimates**

Dep.Vbls	Q	VA
Model	(S7)[a]	(S7)[b]
Labour	0.374***	0.767***
	(0.064)	(0.127)
Materials	0.516***	-
	(0.049)	
$D_1 * Capital$	0.181**	0.284*
	(0.091)	(0.162)
$D_2 * Capital$	-0.078	-0.044
	(0.059)	(0.073)
Cap.Util	0.326***	0.622***
	(0.095)	(0.201)
R&D	0.122**	0.251**
	(0.056)	(0.106)
Ind. Dum.	Inc.	Inc.
Time Dum.	Inc.	Inc.
Period	1990-99	1990-97
N. Obs	11,476	8,687
Sargan T.	83.18	53.12
(df)	(81)	(65)
m1	-10.26	-10.41
m2	0.01	-1.01

* significant at 10% level; ** significatn at 5%; *** significant at 1%.

[a] $IV\dot{s}$: number of workers, physical capital and capacity utilisation lagged levels from $t-2$ to $t-4$; materials lagged levels from $t-2$ to $t-3$ and R&D capital lagged levels in $t-2$; (exogenous variables) growth of capital.

[b] $IV\dot{s}$: number of workers and capacity utilisation lagged levels from $t-2$ to $t-4$; physical capital from t to $t-4$ (as it is considered exgoneous); R&D capital lagged levels from $t-2$ to $t-3$;

Table 4. Temporary Equilibrium 2

Estimation method: **GMM estimates**

Dep.Vbls	Q	VA
Model	(S8)[a]	(S8)[b]
Const.	0.00	0.024
	(0.011)	(0.027)
Labour	0.347***	0.794***
	(0.069)	(0.150)
Materials	0.533***	-
	(0.059)	
Capital	0.076***	0.123***
	(0.025)	(0.041)
R&D	0.134**	0.296**
	(0.058)	(0.155)
Ind. Dum.	Inc.	Inc.
Time Dum.	Inc.	Inc.
Period	1990-99	1990-97
N. Obs	11,476	8,687
Sargan T.	82.97	35.71
(df)	(67)	(39)
m1	-8.69	-10.25
m2	-1.24	-1.74

* significant at 10% level; ** significatn at 5%; *** significant at 1%.

[a]$IV\dot{s}$: number of workers lagged levels from $t-2$ to $t-5$; capacity utilisation from $t-2$ to $t-4$ and materials from $t-2$ to $t-3$; R&D capital lagged levels in $t-2$; growth of "short-run" capital, C*, is used as instrument for stock of capital.

[b]$IV\dot{s}$: number of workers and capacity utilisation lagged levels from $t-2$ to $t-4$; R&D capital lagged levels from $t-2$ to $t-3$; growth of "short-run" capital, C*, is used as instrument for stock of capital.